

# Esta aula

---

## ▶ Plano

- ▶ Variável Proxy
- ▶ Erro de Medida
- ▶ Outliers
- ▶ Variável Dependente Binária

## ▶ Bibliografia

- ▶ Wooldridge, J. M. *Introductory Econometrics: A modern Approach*, 6th Ed.



# Variável Proxy

# Variável Proxy

---

- ▶ Problema: Em algumas situações, não temos dados sobre uma variável importante no modelo.
- ▶ Sabemos que isso levaria a viés na estimação de todos os coeficientes do modelo.
- ▶ Nesse caso, podemos utilizar uma variável proxy, relacionada a variável omitida  $x_3^*$  :

$$x_3^* = \delta_0 + \delta_3 x_3 + v_3,$$

# Variável Proxy

---

- ▶ Se substituirmos  $x_3^*$  por  $x_3$ , temos:

$$y = (\beta_0 + \beta_3 \delta_0) + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \delta_3 x_3 + (u + \beta_3 v_3)$$

- ▶ Os pressupostos necessários para se obter estimativas não-enviesadas de  $\beta_1, \beta_2$
- ▶ 1.  $u$  não deve ser correlacionado com  $x_1, x_2$  e  $x_3^*$
- ▶ 2.  $v_3$  não deve ser correlacionado com  $x_1, x_2$  and  $x_3$

# Erro de Medida

# Erro de Medida

---

- ▶ Erro de Medida na Variável Dependente:  
Suponha que queremos medir  $y^*$  mas observamos apenas  $y$ . Definimos o erro de medida como:
- ▶  $e_0 = y - y^*$
- ▶ Suponha que estimamos:
- ▶  $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u + e_0$

# Erro de Medida

---

- ▶ Suponha que estimamos:
- ▶  $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u + e_0$
- ▶ Quando MQO sera não-enviesado?
- ▶ Apenas se as variáveis explicativas forem não correlacionadas com  $u$  e  $e_0$
- ▶ Porém, se  $E(e_0) \neq 0$ ,  $\beta_0$  será enviesado
- ▶ E a variância do termo de erro será maior levando a estimativas mais imprecisas

# Erro de Medida

---

- ▶ Erro de Medida na Variável Independente:  
Suponha que queremos medir  $x_1^*$  mas observamos apenas  $x_1$ . Definimos o erro de medida como:
- ▶  $e_1 = x_1 - x_1^*$
- ▶ Suponha que estimamos:
- ▶  $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u - \beta_1 e_1$



# Erro de Medida

---

- ▶ Suponha que estimamos:
- ▶  $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u - \beta_1 e_1$
- ▶ Quando MQO sera não-enviesado?
- ▶ Apenas se as variáveis explicativas forem não correlacionadas com  $u$  e  $e_1$
- ▶ Porém, se  $E(e_1) \neq 0$ ,  $\beta_0$  será enviesado
- ▶ E a variância do termo de erro será maior levando a estimativas mais imprecisas

# Erro de Medida

---

- ▶ Se  $\text{Cov}(x_1^*, e_1) = 0$ , então:
- ▶  $\text{Cov}(x_1, e_1) = E(x_1 e_1) = E(x_1^* e_1) + E(e_1^2) = 0 + \sigma_e^2$
- ▶ Nesse caso, MQO será enviesado!



# Outliers

# Outliers

---

- ▶ Reportar as estatísticas descritivas para detectar a presença de outliers.
- ▶ Se a presença de outliers for devido a incorrencia de erro em reportar o verdadeiro valor, pode-se corrigir o valor reportado ou descartar a observação.
- ▶ Se não, deve-se reportar os resultados com e sem a presença de outliers

# Variável Dependente Binária

# Modelo Linear de Probabilidade

---

- ▶ Se a variável dependente for linear, então  $P(y = 1 | x) = E(y|x)$ . Nesse caso:
- ▶  $P(y = 1 | x) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$
- ▶ Nesse caso, cada intercepto beta i mede o impacto de variações marginais em  $x_i$  na probabilidade do evento 1 ocorrer.
- ▶ O valor previsto de  $y$  nesse é a probabilidade do evento 1 ocorrer.

# Modelo Linear de Probabilidade

---

- ▶ Problema: nada impede que o  $y$  previsto não esteja no intervalo  $[0, 1]$ .
- ▶ Geralmente, esse modelo viola também o pressuposto de homocedasticidade.

# Probit e Logit

---

- ▶ Nesse caso, podemos definir:

$$P(y = 1 | \mathbf{x}) = G(\beta_0 + \mathbf{x}\beta)$$

onde  $0 < G(\beta_0 + \mathbf{x}\beta) < 1$



# Probit e Logit

---

- ▶ No modelo probit,  $G(z)$  é a distribuição acumulada da normal padrão  $G(z) = \Phi(z) \equiv \int \phi(v) dv$ , onde  $\phi(z) = (2\pi)^{-1/2} \exp(-z^2/2)$
- ▶ Como o modelo é não-linear, não podemos utilizar OLS
- ▶ Solução: estimação da função de máxima verossimilhança.

# Probit e Logit

---

- ▶ No modelo logit,  $G(z)$  é a distribuição acumulada da variável aleatória logística
- ▶  $G(z) = \exp(z) / [1 + \exp(z)] = \Lambda(z)$

# Probit e Logit

---

- ▶ Em geral, queremos saber o efeito de um incremento de  $x$  em  $P(y = 1 | \mathbf{x})$ ,
- ▶ Nesse caso,  $\partial p / \partial x_j = g(\beta_0 + \mathbf{x}\beta) \beta_j$ , where  $g(z)$  is  $dG/dz$

# Probit e Logit

---

- ▶ Apesar de não poder utilizar o teste F para testar a significância dos coeficientes, podemos utilizar o Likelihood Ratio test:
- ▶  $LR = 2(L_{ur} - L_r) \sim \chi^2_q$

# Probit e Logit

---

- ▶ No modelo linear, calculamos o R-quadrado para medir o grau de explicação do modelo.
- ▶ Em logit/probit, podemos calcular a percentagem corretamente prevista da variável dependente
- ▶ Podemos calcular também o pseudo R<sup>2</sup>, equivalente a  $1 - L_{ur}/L_r$



Obrigada!

